Exercise 1: ROC metrics

Consider a binary classification algorithm that yielded the following results on 10 observations. The table shows true classes and predicted probabilities for class 1:

ID	True class	Prediction
1	0	0.33
2	0	0.27
3	0	0.11
4	1	0.38
5	1	0.17
6	1	0.63
7	1	0.62
8	1	0.33
9	0	0.15
10	0	0.57

a) Create a confusion matrix assuming a threshold of 0.5. Point out which values correspond to true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

b) Calculate: PPV, NPV, TPR, FPR, ACC, MCE and $F1$ measure.

c) Draw the ROC curve and interpret it. Feel free to use `R` or `Python` for the drawing.

d) Calculate the AUC.

e) How would the ROC curve change if you had chosen a different threshold in a)?

Exercise 2: k-NN

a) Let the two-dimensional feature vectors in the following figure be instances of two different classes (triangles and circles). Classify the point (7, 6) – represented by a square in the picture – with a k-NN classifier using $L1$ norm (Manhattan distance):

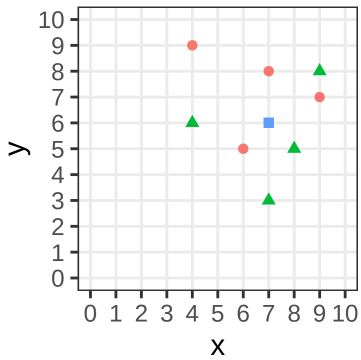
$$d_{\text{Manhattan}}(\mathbf{x}, \tilde{\mathbf{x}}) = \sum_{j=1}^{p} |x_j - \tilde{x}_j|.$$

As a decision rule, use the unweighted number of the individual classes in the k-neighborhood, i.e., assign the point to the class that represents most neighbors.

i) $k = 3$

ii) $k = 5$

iii) $k = 7$



b) Now consider the same constellation but assume a regression problem this time, where the circle-shaped points have a target value of 2 and the triangles have a value of 4.

Again, predict for the square point $(7, 9)$, using both the *unweighted* and the *weighted* mean in the neighborhood (still with Manhattan distance).

i) $k = 3$

ii) $k = 5$

iii) $k = 7$